

Санкт-Петербургский государственный университет
Факультет Прикладной математики – процессов управления
**Кафедра теории систем управления
электрофизической аппаратурой**

Лифантова Екатерина Евгеньевна

Выпускная квалификационная работа бакалавра
**Система поддержки принятия решений при
постановке диагноза**

Направление 010900

«Прикладные математика и физика»

Руководитель научной программы,
доктор физ.-мат. наук,
профессор

Егоров Н. В.

Научный руководитель,
кандидат физ.-мат. наук,
доцент

Гончарова А. Б.

Рецензент,
кандидат физ.-мат. наук,
доцент

Клемешев В. А.

Санкт-Петербург

2016

Содержание

Введение	4
1 Обоснование возможности создания медицинской системы поддержки принятия решения	6
2 Моделирование системы поддержки принятия решения	8
2.1 Контроль и регуляция	8
2.2 Алгоритм постановки диагноза	9
2.3 Теория вероятностей и математическая статистика в диагностике	10
2.4 Обзор существующих экспертных систем диагностики	11
2.5 Обзор представления медицинских данных	12
2.6 Механизм вывода диагностического решения	14
2.7 Проблема неопределенности информации диагностики	15
2.8 Теоретико - вероятностный подход к задаче диагностики	16
3 Программная реализация	19
3.1 Использование оператора множественного выбора	19
3.2 Использование имитации динамических массивов	20
3.3 Применение классов, как объектов представления и обработки данных	21
4 Тестирование системы поддержки принятия решений	24
4.1 Скалярное умножение при тестировании СППР	24
4.2 Применение методов теории вероятностей и математической статистики для решения тестовой задачи	25
Заключение	27

Литература	28
Приложения	31
А База данных	31
В Пример организации выбора наиболее вероятного заболевания с помощью оператора множественного выбора	32
С Описание класса Illnesses	33

Введение

Одной из главных задач врача является правильная постановка диагноза [1]. Современные технологии, внедренные в медицинскую диагностику, позволяют существенно повлиять на постановку предварительного диагноза у пациентов. Существуют различные подходы создания медицинских систем диагностики, например, системы поддержки принятия решений (СППР), которые занимаются неформализованными задачами [2]. Такие задачи нацелены на выявление заболевания пациента. Особенность каждого заболевания обусловлена разным набором симптомов и значениями коэффициентов специфичности. В процессе работы СППР итоговые оценки рассматриваемых заболеваний сравниваются с поступающими факторами проявления симптомов и затем делаются выводы о состоянии области медицинской диагностики.

Автоматизация получения логических выводов применима не только в медицине, но и в других сферах [3]. Так имитационная модель постановки диагноза становится все более обширной темой для изучения и проектирования решений, связанных с различным родом проблем: недостоверностью и малым количеством информации диагностики, аргументацией выдвинутого предположения, а также точностью представленной гипотезы.

В данной работе внимание уделяется именно медицинским экспертным системам, потому что задача поддержки принятия квалифицированного решения по поводу дифференциального диагноза является актуальной. Ускоренный темп жизни приводит к неправильному питанию и стрессовым состояниям, в связи с этим распространены заболевания поджелудочной железы (ПЖ) [4]. Симптомы проявления таких заболеваний различные, зависят от многих факторов, в связи с чем существуют разногласия по поводу диагностики результатов обследований.

Целью работы является создание автоматизированной СППР для постановки диагноза у больных с заболеваниями ПЖ.

Процесс решения неформализованных задач не имеет четко прописанного алгоритма, поэтому приходится применять различные методы и

подходы, которые не всегда приводят к нужному результату.

Для реализации поставленной цели ставятся следующие задачи, позволяющие приблизить процесс автоматизированной диагностики к процессу диагностики осуществляемой экспертом:

- представить медицинские данные в удобном для использования виде;
- разработать программу для нахождения наиболее вероятного заболевания;
- найти способ решения проблемы неполноты входных данных при постановке диагноза;
- рассмотреть способ применения методов теории вероятностей и математической статистики при выявлении наиболее вероятного заболевания;
- проверить предложенный метод решения;
- сделать выводы о проделанной работе.

Возможность проверить адекватность результатов реализованных задач обусловлена наличием анамнезов реальных больных. Обезличенные данные предоставлены врачом - терапевтом Клинической больницы № 10 города Уфа.

1 Обоснование возможности создания медицинской системы поддержки принятия решения

В данной работе производится моделирование системы принятия решений для выявления заболеваний ПЖ. Детальные классификации систем диагностики, области их применения и особенности достаточно подробно изложены Мартемьяновым Ю. Ф. и Лазаревой Т. Я. [2].

С помощью [5], где довольно удобно приведена тщательно проработанная систематика болезней по анатомическим особенностям человека, были проанализированы болезни, связанные с нарушением в работе ПЖ. Четкое представление данных способствовало лучшему представлению об изучаемой области.

Для создания СППР были рассмотрены общие подходы создания процедур принятия решений [1]. Изучена математическая постановка задачи для автоматизированных систем, а также проанализирован метод экспертов, который хорошо применим для реализации медицинских систем диагностики [2]. В методе экспертов была представлена идея субъективного оценивания коэффициента специфичности, отвечающего за некоторый параметр системы.

Теории вероятностей и математической статистики при моделировании постановки диагноза представлены в [9]. Процесс диагностики врачом — это сложный мыслительный процесс, а его описание представляется труднореализуемым для программного моделирования. В связи с этим, в работе представлены разные варианты изучаемой модели экспертных систем, в которых применяются свои особенности и упрощения, в зависимости от конкретных принципов формализации процессов диагностики. Применение методов теории вероятностей обусловлено попытками приблизить процесс математической постановки диагноза, к процессу постановки диагноза врачом.

Существование разных подходов к процессу диагностики хорошо по-

казывает наличие ряда статей. Например, в [6] вычислительная диагностика основана только на применении формально - вероятностной методики. В работе [7] представлена модель постановки диагноза с помощью имитационного моделирования и дерева решений. Возможность постановки диагноза при помощи корреляционного и дискриминантного анализа представлена в статье [8]. В ходе изучения предложенных статей возникла идея представить новый способ обработки и анализа данных, не встречающийся ранее.

2 Моделирование системы поддержки принятия решения

2.1 Контроль и регуляция

Группа операций управления формирует процесс управления, таким образом, управление представляет из себя совокупность процессов, обеспечивающих достижение определенной цели. Множество технических средств, выполняющих операции управления без непосредственного участия человека, является объектом управления автоматизированного устройства.

Частным случаем управления является регулирование [10]. Контроль и регуляция подразумевает под собой некоторые внешние изменения контролируемого параметра (КП). Из-за того, что КП и внешние изменения подвергаются воздействию совокупности факторов, то постоянное изменение, сравнение с соответственной постоянной — точкой контроля, и корректировка любых отклонений позволяют КП удерживать на постоянном уровне. Данный замкнутый тип управления называется системой контроля по принципу обратной связи, или замкнутым контролем (рис. 1).



Рис. 1. Система регулирования с обратной связью: принцип.

Одна из прикладных задач данного вида управления — описание состояния изучаемой области, вышедшей из положения равновесия.

Анализируя результаты по классу первичной заболеваемости органов пищеварения в России в период с 2010 по 2014 год, можно выявить тенденцию роста числа больных [11], что показывает актуальность задачи постановки диагноза с заболеванием ПЖ.

Определение 2.1 Поджелудочная железа [10] (*лат. pancreas*) — железа пищеварительной системы, вырабатывающая панкреатический сок и обладающая одновременно эндокринной функцией.

2.2 Алгоритм постановки диагноза

В настоящее время диагностический процесс, производимый врачом, представляет из себя соотнесение дефектных показателей (симптомы заболеваний) с различными органами, тканями, а также данными лабораторных анализов. Внедрение современных технологий позволило повлиять на постановку точного диагноза, а также качественнее отслеживать и оценивать различные риски, влияющие на процесс развития заболевания.

Результирующим действием диагностики должны быть постановка правильного диагноза и выписка рекомендованного курса лечения. Формализация процесса диагностики представлена на рис. 2.

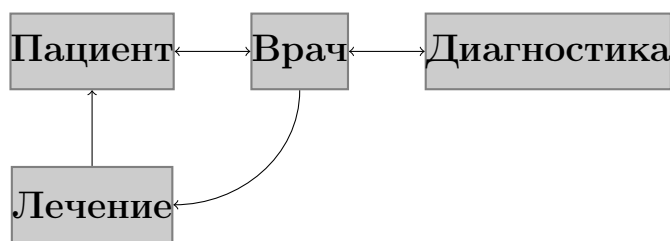


Рис. 2. Алгоритм диагностики.

СППР позволяют врачу не только проверить выдвинутую гипотезу в пользу выбранного заболевания, но и обратиться за советом в трудных для принятия решения ситуациях. СППР должны способствовать более раннему диагностированию подозрения на заболевание у врачей - терапевтов или врачей общей практики.

2.3 Теория вероятностей и математическая статистика в диагностике

Исследования в области медицины с применением теории вероятностей и математической статистики основаны на экспериментальных данных. Методы математической статистики используют для решения следующих проблем, связанных с постановкой диагноза:

- возможности описания статистики;
- сравнения выборок;
- исследования зависимостей между переменными;
- сужения пространства признаков;
- классификации:
 - с учителем (дискриминантный анализ);
 - без учителя (кластерный анализ и группировка).

В дискриминантном анализе, успешно применяемом для диагностики заболеваний в медицине, обучающими выборками, элементы которых принадлежат к известным классам, являются, например, классы «больной/здоровый». Вновь появляющийся объект необходимо отнести к одному из этих классов, поэтому термин «диагностика» в рамках изучаемого подхода может рассматриваться как синоним «дискриминации».

Кластеризация и группировка предназначены для выявления и выделения классов, в отличие от дискриминантного анализа, где классы заданы. Теория математической статистики применительно к медицине описана в [9].

Численный способ описания того, что происходит на практике при постановке диагноза очень хорошо описывает байесовский метод — один из статистических методов исследования вероятностей проявления отдельных

симптомов при отдельных заболеваниях, позволяющий вычислить вероятность того, что данный больной страдает определенным заболеванием, относящимся к рассматриваемой группе. Применение формул вероятностей позволяют вычислить условную вероятность заболевания при данном наборе симптомов на основании априорной вероятности заболевания и условной вероятности наблюдения симптомов.

Данный метод диагностики имеет ряд проблем, связанных со следующими факторами:

- большим объемом информации (организация, размещение и хранение данных);
- количественным характером информации (нельзя решить задачи качественного характера);
- пренебрежением взаимодействия различных областей диагностики.

2.4 Обзор существующих экспертных систем диагностики

На сегодняшний день известно большое количество медицинских систем диагностики. Одними из первых в этой области следует выделить такие системы как MYCIN и PUFF.

MYCIN была спроектирована в середине 70-х годов для диагностирования бактерий, которые вызывают тяжелые инфекции (бактериемия и менингит), и для указания рецепта антибиотиков в зависимости от массы тела пациента. Механизм работы в MYCIN представляет собой первоначальный опрос пациента, прямой вывод с использованием некоторых правил нечеткой логики и обратный вывод. В настоящий момент MYCIN используется в основном для обучения медицинских работников, а ее механизм логического вывода был задействован при разработке таких экспертных систем, как NEOMYCIN и PUFF для исследования легочных заболеваний [12].

Система PUFF диагностирует наличие и степень тяжести заболева-

ния легких у пациента, моделируя измерения параметров дыхания в лаборатории функций легких. Данная экспертная система основана на правилах с обратной цепочкой рассуждений [13].

С развитием компьютерных технологий разрабатываются системы, способные делать логические выводы и принимать решения. Например, суперкомпьютер IBM Watson способен не только ставить диагнозы, но и определять наиболее оптимальный курс лечения. Российская разработка AImedica — еще одна эффективная система для помощи врачу, основанная на базе знаний, содержащихся в более чем 22 млн медицинских источников, интегрирующая накопленные человечеством знания в области медицины и представляющая их в удобном справочном формате. Интеллектуальная система AImedica, так же, как и IBM Watson, оснащена вопросно - ответной формой искусственного интеллекта, что позволяет вести диалог с пользователем, направленный на постановку диагноза за минимальное количество вопросов о симптомах. Более подробное описание приведенных систем можно найти в литературе [13, 14].

Несмотря на то, что медицинские экспертные системы созданы и уже тестируются, система диагностики не может нести какую - либо ответственность за принятое решение. Задача любой системы — проконсультировать врача в определенной области или определить возможного специалиста, к которому следует обратиться, на основании полученных данных.

2.5 Обзор представления медицинских данных

От выбора вида представления базы знаний в медицинской системе зависит выбор метода представления решения [15]. Поэтому при разработке СППР важно учитывать ограничения предметной области и возможность представления процедурных знаний [16].

В связи с этим предлагается имитационная модель рассматриваемой задачи диагностики заболеваний ПЖ, в которой симптомокомплекс представляется в виде пары «ключ-значение» или «объект-ключ-значение». Состояние области диагностики заболеваний представлено в виде таблицы 1.

Столбцы задают наименование болезни, а строки — симптомы, диагностируемые у больных с нарушением работы ПЖ. В ячейках пересечения «симптом-болезнь» ставится численное значение, свидетельствующее об его наличии или отсутствии.

Симптокомплекс состоит из шести заболеваний, вызванных нарушением в работе ПЖ (см. приложение А):

- муковисцидоз;
- макроамилаземия;
- злокачественные опухоли ПЖ;
- острый панкреатит;
- хронический панкреатит;
- псевдокисты ПЖ.

Таблица 1. Пример базы данных [5, 18–21].

Симптомы \ Заболевания	Карциномы ПЖ	Панкреатит	Хронический панкреатит
тошнота	1	1	1
рвота	1	1	1
метеоризм	0	1	0
общая слабость	1	1	0
быстрая утомляемость	1	0	0
снижение веса	0	0	1
высокая температура	1	1	0

Особенность каждого заболевания обусловлена разным набором симптомов и значением коэффициентов специфичности каждого из них. По итоговым оценкам рассматриваемых заболеваний, сравниваемых с поступающими факторами проявления симптомов, можно сделать выводы о состоянии области медицинской диагностики.

На основе анализа представления данных, для отображения статистических значений о текущем состоянии области диагностики используется вектор, который формируется из значений столбца заболевания. Таким образом, можно утверждать, что каждому рассматриваемому заболеванию соответствует свой вектор-столбец.

2.6 Механизм вывода диагностического решения

При построении пользовательского вектора I важно учитывать, что сам по себе этот объект не представляет никакого значения при отсутствии дополнительной информации о данном векторе. В качестве обработки входного вектора предлагается его скалярное умножение с векторами заболеваний, в результате чего образуется новый параметр системы данных — ассоциативный массив скалярного перемножения векторов заболеваний и пользовательского вектора T . Для возможных дальнейших обращений необходимо инициализировать данные такими параметрами как: имя, фамилия, возраст и пол:

$$< Name, Surname, Age, Sex, I, T >$$

При построении вышеуказанной модели учитывались необходимые данные для комплектования базы данных, однако для возможных дальнейших усложнений системы важно учитывать, что такого рода представления предметной области может оказаться недостаточным и потребуются дополнительные усовершенствования.

Особенностями данной модели представления данных является простота и легкость управления, способность наглядного отображения организации области диагностики, а также эффективность при дальнейшем обращении к значению вектора.

2.7 Проблема неопределенности информации диагностики

В медицинской практике часто встречается неопределенность информации диагностики. Неопределенность становится большой проблемой на пути к постановке точного диагноза, так как мешает выбрать лучшее решение, и, следовательно, может стать причиной некачественного ответа. В данной работе под неопределенностью понимается недостоверность наличия симптома, то есть субъективность оценивания специфичности симптома в симптомокомплексе.

Экспертная система диагностики должна корректно функционировать даже при условиях неполной определенности. Из-за отсутствия некоторых входных данных система не должна останавливаться на полученном решении, которое может быть неточным. Всегда существует возможность введения любых правил в систему, способную работать с неполной информацией; тогда результат работы программы будет зависеть от частоты проявления определенных характеристик, если не было введено пороговое значение.

Формально модель с неполной информацией может быть представлена в следующем виде (рис. 3).

Компонента пользовательского вектора, создающегося в результате выполнения программы, принимает значение 1, если симптом точно присутствует, 0 — отсутствует и X, если пользователь сомневается. Под X подразумевается, что ячейка вектора заполняется 0 или 1. Теперь количество рассматриваемых векторов возрастет до 2^n , где n — число позиций с неопределенными значениями.

В условиях неопределенности при постановке диагноза следует вводить различные правила, учитывающие особенности конкретной болезни при имеющихся у больного симптомах. Эти особенности могут быть выражены с помощью присвоения веса каждому симптому в зависимости от его значимости в иерархии симптомокомплекса. Подобная методика поможет отобразить знания не только в виде «наличие/отсутствие», но и

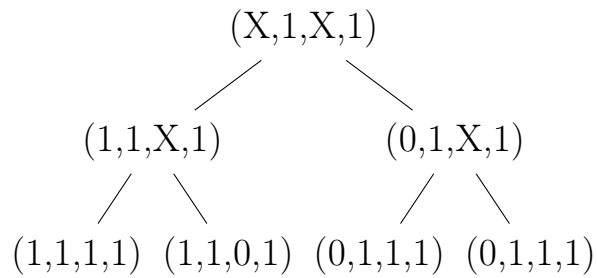


Рис. 3. Пример представления решения проблемы неопределенности информации диагностики.

учесть вклад каждого симптома при дальнейшей обработке пользовательского вектора. Функция присваивания веса представляет собой субъективную меру. Под субъективной мерой понимается релевантность элемента понятию, определенному нечеткой совокупностью. Основываясь на базе знаний и введенных фактах проявления симптомов можно определить количественные оценки каждого заболевания.

2.8 Теоретико - вероятностный подход к задаче диагностики

Рассмотрим ограниченную группу d различных заболеваний (D_1, D_2, \dots, D_d) , при этом каждый больной страдает только одним из них. Также имеется список s симптомов, на основе которых можно получить необходимую информацию, обозначаемых (S_1, S_2, \dots, S_s) . Считаем, что все симптомы дискретны, т.е. каждый из них относится к одному из двух классов: «наличие» или «отсутствие».

Учитывая накопленные знания по рассматриваемому набору заболеваний, можно задать для каждой болезни D_j вероятность $P(S|D_j)$ наблюдения определенного симптома S . До начала обследования представляется возможным посчитать *априорную вероятность* $P(D_j)$ того, что больной страдает одним из рассматриваемых заболеваний. Однако, важно принять во внимание, что вероятность встречи распространенной болезни выше чем заболевания, встречающегося крайне редко. В связи с этим можно сделать вывод о различных значениях априорных вероятностей для

заданного множества D .

Далее при диагностике выявляется список наиболее вероятных заболеваний, которые характеризуют определенный симптокомплекс S . Таким образом, для быстрой постановки предварительного диагноза врачу необходимо знать вероятность каждого предполагаемого заболевания *при данном* симптокомплексе, т. е. $P(D_j|S)$. В нашем случае *апостериорная вероятность* $P(D_j|S)$ пропорциональна произведению априорной $P(D_j)$ на функцию правдоподобия $P(S|D_j)$:

$$P(D_j|S) \sim P(D_j) P(S|D_j) \quad (1)$$

Т.к. сумма всех апостериорных вероятностей для различных заболеваний должна быть равна 1, то (1) запишется в следующем виде

$$P(D_j|S) = \frac{P(D_j) P(S|D_j)}{\sum_{j=1}^d P(D_j) P(S|D_j)}, \quad j = \overline{1, d} \quad (2)$$

Формула (2) позволяет вычислить вероятность справедливости выдвинутой гипотезы на основании ее априорной вероятности и некоторых эмпирических данных. Этот метод можно использовать с полным оправданием, только если альтернативные заболевания имеют априорные вероятности.

При вычислении апостериорной вероятности важно с осторожностью относиться к вычислению $P(S|D_j)$, т.к. база знаний, на основе которой вычисляется функция правдоподобия, может содержать мало данных о частоте обнаружения определенного симптокомплекса S , несмотря на емкость материала, так как это может зависеть от района, его экологии и многих других факторов. Если есть s различных симптомов, которые являются дихотомичными, то число различных симптокомплексов будет равно 2^s . При $s > 20$ применение метода оказывается невозможным из-за большого значения числа ($> 10^6$) возможных симптокомплексов. В связи с этим появляется необходимость несколько упростить предположение. Одним из допущений может стать предположение о независимом распределении отдельных симптомов. Если принять, что s симптомов статистически неза-

висимы, тогда соотношение

$$P(S|D_j) = \prod_{i=1}^s P(S_i|D_j), \quad j = \overline{1, d} \quad (3)$$

позволяет на основе имеющегося материала с достаточно высокой точностью вычислить вероятность $P(S_i|D_j)$.

Перейдем к программной реализации разработанных заданных моделей.

3 Программная реализация

3.1 Использование оператора множественного выбора

Заключение об адекватности построенной модели истинно, если истинны все заключения об адекватности компонент модели. Имеется в виду, что данная система диагностики должна функционировать и при малых количествах симптомов.

Таблица 2. Симптомы и заболевания.

	v_1	v_2	v_3	v_4
u_1	1	0	1	0
u_2	1	1	0	1
u_3	1	1	1	1
u_4	0	1	1	1

Прежде чем приступать к непосредственной работе с созданной базой знаний, рассматривается упрощенная модель. Пусть имеется таблица 2, в которой по строкам откладываются симптомы $u_i, i = \overline{1, 4}$, а по столбцам заболевания $v_j, j = \overline{1, 4}$. В данном примере соответствие симптомов и заболеваний отмечено произвольным образом. В нашем случае 0 соответствует отсутствию симптома, 1 — его наличию. Таким образом, получается матрица, состоящая из векторов-симптомов.

То есть в матричном виде ее можно представить как матрицу U

$$U = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

Теперь на языке C++ в среде программирования Visual Studio 2012 описывается выбор наиболее вероятного заболевания посредством сравнения пользовательского вектора, создающегося в результате выполнения программы, и заданных векторов. Все последовательные операции сравнения производятся с помощью оператора множественного выбора `switch()` (см.

приложение В). Одновременно в ходе выполнения программы ведется учет количества совпадений с помощью кумулятивной переменной s , которой прибавляется значение $+1$ если имеется совпадение в позиции у пользовательского вектора и вектора-симптомов. В дальнейшем, оперируя количеством совпадений, можно сделать вывод о наиболее вероятном заболевании.

Программа показала хороший результат для матрицы размерности 4×4 . В результате разработки кода, обнаружился ряд сложностей, касающихся именно реализации метода. Одним из самых главных недостатков оказалось ограничение метода относительно размерности используемой матрицы.

3.2 Использование имитации динамических массивов

Для реализации представления объектов области диагностики заболевания в виде векторов потребовалось использовать массивы более высокого уровня, например, последовательный контейнер `vector` и ассоциативный контейнер `map` из стандартной библиотеки шаблонов STL. Ассоциативный контейнер позволяет представить базу знаний в виде «ключ-значение», или в случае рассматриваемой задачи «заболевание-вектор симптомов». Кроме того, применение контейнера позволяет достаточно легко сделать возможным поиск и упорядочивание по ключу. `Vector` дает возможность использовать встроенные функции для описания присваивания значения или поиска элемента, что удобно при большом объеме значений.

Таким образом, в процессе выполнения программы на первом этапе будет производиться заполнение данных, связанных с симптомами. На следующем этапе при помощи скалярного произведения вектора - симптомов пациента и матрицы U из векторов - симптомов заболеваний будет вычисляться специальная характеристика. В итоге каждому обратившемуся пациенту в соответствие будет ставиться набор параметров: вектор симптомов, вектор-результат скалярного умножения, а так же функция вывода наиболее вероятной болезни. Поэтому целесообразно оснастить программу

новым типом данных, включающего в себя этот набор параметров, например, struct DB.

```
struct DB
{
    vector<int> symptomy;
    vector<int> result;
    int vivod();
};
```

Внедрение новых моделей представления данных в код программы расширили возможности использования, упростив код и сделав его более гибким к изменению значений элементов векторов.

3.3 Применение классов, как объектов представления и обработки данных

В данной работе было произведено усовершенствование метода диагностики, которое подразумевает внедрение не только новых библиотек и функций, но и нового способа алгоритмического и программного представления. Для принятия решения необходима предварительная оценка ситуации, ее параметров, затрат и т.п. Создавая систему поддержки принятия решения, важно учитывать эти факты.

Классы широко применяются в объектно - ориентированном программировании для описания различных процессов со сложными и многократными обращениями. Классы в C++ — совокупность свойств и методов, описывающих некоторые объекты, которые однозначно определяют экземпляр класса, если задано их множество и свойства. Ранее предлагалось использовать встроенные контейнеры для определения области диагностики, которые давали достаточное представление о болезни, но не полное.

При создании класса Illnesses (см. приложение С) учитывались особенности изучаемой области. В данной работе расширена сфера применения методов контейнеров, так как применялись они для матриц, состоящих не только из 0 и 1, а также для взвешенных позиций матрицы. Взвешенной позиции матрицы приписывается численное значение, свидетельствующее о роли симптома в иерархии симптомокомплекса. Применение весов было вы-

брано для решения проблемы неопределенности входных данных пациента. Веса были предоставлены врачом - экспертом клинической больницы № 10 города Уфа Ерохиной Т. А.

Работу класса Illnesses можно описать следующими этапами:

На первом этапе в конструкторе осуществляется считывание названия текстового документа, в котором хранятся значения векторов - симптомов. При этом, происходит проверка на допустимость имени файла, проверяются права доступа к файлу и его существование.

Второй этап включает в себя считывание файла, заполнение названий болезней, заполнение векторов - заболеваний и создание ассоциативного массива mapIllnesses, хранящий в себе характеристики болезней.

Третий этап представляет из себя обработку пользовательского вектора и векторов - заболеваний, которая описана в функциях bayes и scalmultiplay, принимающие в качестве входного параметра пользовательский вектор. Также одним из входных параметров данных функции scalmultiplay является некоторое пороговое значение adge, которое принимается таким образом, чтобы минимизировать затраты на рассмотрение маловероятных симптомов.

Для хранения данных спроектирован класс DB, оснащенный такими полями как Name, Surname, Age, Sex и функцией create(), которая отвечает за создание текстового документа, в котором хранятся данные обратившегося пациента, результаты выполнения функций bayes и scalmultiplay.

Особенностью реализации классов для обработки и хранения входных данных является то, что с помощью их внедрения в тело программы объекту класса ставится набор параметров, представляющих его полную характеристику.

Априорные вероятности каждого заболевания представляют из себя численное значение, которое является степенью проявления болезней в клинической практике (данные предоставлены врачом-терапевтом Ерохиной Т. А. (таб. 3).

Таблица 3. Априорные вероятности проявления болезней.

Наименование болезни	$P(D_j)$
острый панкреатит	0.394
хронический панкреатит	0.315
злокачественные опухоли ПЖ	0.156
псевдокисты ПЖ	0.079
муковисцидоз	0.04
макроамилаземия	0.016

4 Тестирование системы поддержки принятия решений

В качестве апробации созданной программы используются данные обезличенных анамнезов предоставленные врачом в количестве двенадцати экземпляров. Процесс верификации программы представляет из себя выявление наиболее вероятного заболевания посредством двух операций: скалярного перемножения и с помощью теории вероятности, то есть байесовского метода.

4.1 Скалярное умножение при тестировании СППР

Скалярное произведение представляет из себя численный результат перемножения двух векторов. В случае рассматриваемой задачи этот метод стоит несколько усложнить. Неопределенность информации может отложить свой отпечаток на конечный результат. Для выявления некоторой зависимости от входных данных были рассмотрены следующие виды скалярных произведений:

1. суммируются только те позиции векторов заболеваний, которым соответствует в пользовательском векторе 1;
2. так как -1 выражает неопределенность пользователя в наличии симптома, то и расчет следует рассмотреть для неопределенной позиции вектора с половинным вкладом в результат итоговой суммы;
3. принять во внимание полный вклад неопределенной позиции в итоговый результат.

Проведя тестовые запуски программы для трех рассматриваемых видов расчета скалярного произведения были получены следующие результаты.

Наилучшим результатом считается результат, показавший наибольшее количество совпадений. Лучшие показатели точности были выявлены

для случая 3 (9 совпадений из 12 и 2 больных с одинаковыми численными значениями для двух болезней, одно из которых действительно диагностируемое), в то время как худший результат показал случай 1, где не обнаружилось ни одно совпадения, кроме случаев одинаково значащих болезней, одно из которых реально диагностируемое.

В связи с тем, что в структура симптома не является однородной по отношению к позициям векторов-заболеваний, возникает необходимость убрать из рассмотрения наименее вероятные симптомы с помощью порогового значения. После проверки на эффективность оптимальным оказалось значение порога, равное 0.4. Оно было выбрано с помощью варьирования значений в следующих пределах $[0.0, 0.7]$. При увеличении порогового значения наблюдалась тенденция к увеличению качества ответа от 0 до 0.7, что проявлялось в виде наибольшего числа совпадений диагностируемых больных у системы и у врача - специалиста. Это тенденция наблюдалась до значения 0.4, после чего увеличения качества не наблюдалось. Выбор предела исследования в качестве 0.7 обусловлен соображениями, что при еще большем увеличении могут быть отсеяны и важные симптомы, имеющие значимый вес в симптомакомплексе.

4.2 Применение методов теории вероятностей и математической статистики для решения тестовой задачи

В связи с малым объемом изучаемой области, состоящей из $d = 6$ заболеваний (D_1, \dots, D_6) и $s = 45$ симптомов (S_1, S_2, \dots, S_{45}), применение метода, описанного в 2.8 оказывается некачественным. Это связано также с тем, что если есть хотя бы одно несовпадение пользовательского вектора и векторов заболеваний, а такое может произойти, то вся апостериорная вероятность будет равна нулю в силу перемножения априорной вероятности и вероятности обнаружения симптома при конкретной болезни.

При применении теории вероятности для автоматизированной диагностики в случае рассматриваемой задачи происходит перемножение толь-

ко тех значений пользовательского вектора и векторов заболеваний, которые отличны от нуля.

В случае простой дихотомии симптомы, в которых пациент сомневается, принимаются равными нулю, так как для более сложного случая рассматриваемый метод не применим.

Заключение

В результате работы была разработана и реализована система поддержки принятия решений. При этом были решены следующие задачи:

- была представлена база знаний заболеваний ПЖ в виде таблицы, где каждой болезни соответствовал свой вектор столбец. Позиции векторов принимают значения от 0 до 1 в зависимости от их значимости в симптомокомплексе;
- разработаны и опробованы различные подходы к созданию медицинской СППР, базируясь на применении оператора множественного выбора и встроенных контейнеров;
- рассмотрена проблема неполноты входных данных, а также способ ее решения с помощью присвоения веса симптомам;
- программно реализован поиск наиболее вероятного заболевания с помощью двух методов: скалярного перемножения векторов и байесовского подхода;
- разработанный программный комплекс был опробован на данных анамнезов обезличенных пациентов, где были прописаны симптомы и диагностируемые заболевания. Байесовский метод оказался плохо применим для выборки малой размерности, поэтому был разработан новый способ обработки данных, основанный на понятии скалярного произведения. Байесовский метод уступил скалярному произведению в точности 25% против 83,3%;

Внедрение нового метода скалярного умножения в программную реализацию СППР позволили рассмотреть новый способ определения наиболее вероятного заболевания и проверить его эффективность по сравнению с уже известным методом.

Список литературы

- [1] Постовит Ю. Ф. Диагноз и диагностика в клинической медицине. Учеб. пособие. Л.: ЛПМИ, 1991. 96 с.
- [2] Мартемьянов Ю. Ф., Лазарева Т. Я. Экспертные методы принятия решений. Учеб. пособие. Тамбов: Изд-во Тамб. гос. тех. ун-та, 2010. 80 с.
- [3] Телегина М. В., Янников И. М. Программная оболочка для создания экспертных систем Grafexpert // Известия Южного федерального университета. Технические науки. 2014. Вып. 16. С. 67–73.
- [4] Пельц В. А. Современное состояние диагностики хирургического лечения острого панкреатита // Сибирский медицинский журнал. 2010. Т. 25, вып. 4. С. 27–32.
- [5] Wallach J. B. Interpretation of Diagnostic tests / ed. by LIPPICONT WILLIAMS&WILKINS. Philadelphia, Pa: Wolters Kluwer business. 2007. 1026 p.
- [6] Гриф М. Г., Юмчмаа А. Применение экспертных систем пульсовой диагностики // Сборник научных трудов НГТУ. 2015. Вып. 3. С. 114–133.
- [7] Сумина Ю. Е., Львович И. Я. Разработки логических моделей постановки диагноза заболеваний молочных желез с помощью имитационного моделирования дерева решений // Вестник Воронежского государственного технического университета. 2010. Т. 6, вып. 10. С. 41–50.
- [8] Коровин Е. Н., Нехаенко Н. Е., Юрьева К. А. Математическое моделирование процесса постановки диагноза «хронический аднексит» на основе корреляционного и дискриминантного анализа // Вестник Воронежского государственного технического университета. 2012. Т. 8, вып. 9. С. 47–49.
- [9] Бейли Н. Математика в биологии и медицине. М.: Мир, 1970. 327 с.
- [10] Зильбернагель С., Деспопулос А. Наглядная физиология. М.: БИНОМ. Лаборатория знаний, 2013. 408 с.

- [11] Суринов А. Е., Баранов Э. Ф., Бугакова Н. С., Гельвановский М. И., Гохберг Л. М., Дианов М. А., Елизаров В. В., Житков В. Б., Забелин В. Н., Иванов Ю. Н., Кевеш А. Л., Косарев А. Е., Лайкам К. Э., Масакова И. Д., Нестеров В. Н., Оксенойт Г. К., Рыбак О. П., Рябушкин Б. Т., Татаринов А. А., Хорошилов А. В. Российский статистический ежегодник 2015. М: 2015. 728 с.
- [12] Журнал «Компьютерные вести On-line». Вып. 40, 2005 г. «Эволюция экспертных систем. История и перспективы» [Электронный ресурс]: URL:<http://tpl-it.wikispaces.com/MYCIN> (дата обращения: 15.05.2014).
- [13] Aikins J. S., Kunz J. C., Shortliffe E. H. PUFF: an expert system for interpretation of pulmonary function data. Computers and Biomedical Research, vol. 16, 1983. P. 199–208.
- [14] Басманов С. Н., Басманова А. А. Обзор эволюции экспертных систем в медицине с точки зрения соответствия основным признакам // Перспективы развития информационных технологий. 2014. Вып. 21. С. 126–130.
- [15] Ле Н. В., Камаев В. А., Панченко Д. П., Трушкина О. А. Модель представления знаний при создании медицинской экспертной системы дифференциальной диагностики // Известия Волгоградского государственного технического университета. 2014. Т. 20, вып. 6. С. 41–50.
- [16] Ле Н. В., Камаев В. А., Панченко Д. П., Трушкина О. А. Обзор подходов к проектированию медицинской системы дифференциальной диагностики // Известия Волгоградского государственного технического университета. 2014. Т. 20, вып. 6. С. 50–58.
- [17] Aimedica [Электронный ресурс]: URL:<https://aimedica.ru> (дата обращения: 11.04.2015).
- [18] LookMedBook [Электронный ресурс]: URL:<http://lookmedbook.ru> (дата обращения: 10.03.2015).

- [19] ProAnalizu.ru [Электронный ресурс]: URL:<http://pro-analizu.ru> (дата обращения: 17.03.2015).
- [20] Helix [Электронный ресурс]: URL:<http://helix.ru> (дата обращения: 15.05.2015).
- [21] Данилова Л. А. Анализ крови и мочи. Изд. 4-е, исправ. СПб.: Салит-Медкнига, 2003. 128 с.
- [22] Пытьев Ю. П., Газарян В. А., Росницкий П. Б. Сравнительный анализ эффективности вероятностного и возможностного алгоритмов медицинской диагностики // Вестн. Моск. ун-та. Серия 3: Физика, астрономия. 2014. Вып. 3. С. 8–14.

Приложения

А База данных

Типы	Симптомы	Муковисцидоз	Макромилаземия	Карциномы поджелудочной железы	Панкреатит	Хронический панкреатит	Псевдокисты поджелудочной железы
Болезни	Область расположения						
	верхняя часть живота	0	0	0	0.1	0.1	0
	эпигастральная область	0	0	0.9	0	0	0
	Какие						
	интенсивные	0	0	1	0	0	0
	постоянные	0	0	1	0	0	0.1
	мучительные	0	0	1	0	0	0
	сильные	0	0	0	0.9	0	0
	невыносимые	0	0	0	0.5	0	0
	глубокая	0	0	0	0	0	0.1
	Усиливаются						
	после еды	0	0	0.9	0	0	0
	по ночам	0	0	0.9	0	0	0
	в положении лежа на спине	0	0	0.8	0	0	0
	после приема жирной пищи	0	0	0	0	0.8	0
	после приема алкоголя	0	0	0	0	0.9	0
	Характер боли						
	распирающий	0	0	0	0.8	0	0
	жгучий	0	0	0	0.9	0	0
	Отдает						
	в поясницу	0	0	0	0.8	0	0
Диспепсия	тошнота	0	0	0.6	0.6	0.5	0
	рвота	0	0	0.7	0.6	0.4	0
	расстройство аппетита	0	0	0.7	0.7	0	0
	эпигастральное вздутие	0	0	0.8	0.6	0	0
	запоры	0	0	0.4	0	0	0.4
	отрыжка	0	0	0.1	0.1	0.1	0
	изжога	0	0	0.1	0.1	0.1	0
	ощущение неприятного вкуса во рту	0	0	0.1	0.1	0.1	0
	жидкий стул	0	0	0	0	0.4	0.4
	метеоризм	0	0	0	0.5	0	0
Фонетические	Механическая желтуха	0	0	0.6	0	0	0
	Сахарный диабет	0	0	0.6	0	0	0
	Асцит	0	0	0.9	0	0	0
	Общая слабость	0	0	0.9	0.7	0	0
	Быстрая утомляемость	0	0	1	0	0	0
	потеря интереса к привычной работе	0	0	0.9	0	0	0
	депрессия	0	0	0.9	0	0	0
	психическая заторможенность	0	0	0.9	0	0	0
	головные боли	0	0	1	0	0	0
	головокружения	0	0	1	0	0	0
Фонетические	нарушение сна	0	0	0.9	0	0	0
	цианоз и бледность	0	0	0.9	0	0	0
	кожных покровов	0	0	0.9	0	0	0
	сухость слизистых оболочек рта, носа, глаз	0	0	0.4	0	0	0
	повышение температура тела	0	0	0.7	0.4	0	0
	повышенная потливость	0.1	0	0.7	0	0	0
	различные типы анемии	0	0	0.9	0	0	0
	снижение иммунитета	0	0	1	0	0	0
	Снижение веса	0	0	0	0	0.7	0
	Вздутие живота	0.1	0	0	0.2	0.2	0

В Пример организации выбора наиболее вероятного заболевания с помощью оператора множественного выбора

```
...
cout<<"Наблюдаются ли у больного симптом v1?"<<endl;
cin>>k;
int s=0;
switch (k)
{
    case 0:// не наблюдается 1й симптом
        cout<<"Наблюдаются ли у больного симптом v2?"<<endl;
        cin>>f;
        switch (f)
        {
            case 1:// не наблюдаются 1й и 2й симптомы, но наблюдается 3й
                s=s+1;
                cout<<"Возможные заболевания с симптомом v3:"<<endl;
                for (int i=0; i<4; i++)
                {
                    if (p[i][2]==1)
                        cout<<"u"<<i+1<<endl;
                    else
                        cout<<" ";
                }
                cout<<"s="<<s<<endl;
                cout<<"Наблюдаются ли у больного симптом v4?"<<endl;
                cin>>y;
            ...
        }
    }
}
```


С Описание класса Illnesses

```
#pragma once
#include <vector>
#include <map>
#include <string>
using namespace std;

// Считывает, хранит и обрабатывает болезни и симптомы
class Illnesses
{
private:
    int illnessCount;
    int symptomCount;

    // Название болезни – список значений симптомов
    map<string, vector<double>> mapIllnesses;

    // Название болезни – список значений симптомов, подходящих к
    userSymptoms
    map<string, vector<double>> suitableIllnesses;

    // Список значений симптомов пользовательского вектора
    vector<double> userSymptoms;

    map<string, double> aprior;
    map<string, double> pravdopod;
    map<string, double> aposterior;

public:
    Illnesses(string address);
    ~Illnesses(void);

    // Название болезни – список значений симптомов, подходящих к userSymptoms
    map<string, vector<double>> getSuitableIllnesses();

    // Ищет болезни, соответствующие симптомам в dataDict
    // <returns>true – подходящие болезни найдены и содержатся в
    suitableIllnesses
    bool foundingSuitableIllnesses(vector<double> userVector);

    // Вычисляет нормы normNumber для подошедших болезней, с учетом порогового
    значения edge
    map<string, double> getNorms(int normNumber, double edge);

    void printUserSymptoms();
    int getSymptomCount();
    map<string, double> bayes(vector<double> userVector);
    map<string, double> scalmultiply(vector<double> userVector);

private:
    map<string, vector<double>> readIllnesses(string address);

    // Проверяет соответствие количества симптомов в dataDict, а также
    корректность значений в нем
    bool isCorrectData(vector<double> userVector);

    // Квадрат, сумма достоверных (1) симптомов минус сумма только тех
    недостоверных (-1) симптомов, которые меньше edge
    double differentSubNorm(double edge, vector<double> data, vector<double>
    userVector);
};
```